EPJ Web of Conferences **60**, 20055 (2013)
DOI: 10.1051/epjconf/20136020055

The ZZ/WZ signal as a background to ZH/WH ($\ell\ell$,$\nu\nu$)/$\ell\nu b\bar{b}$ searches with the CMS experiment.

Caterina Vernieri[1,2,a]

[1] *Scuola Normale Superiore, Pisa, Italy*
[2] *INFN, Sez. Pisa, Italy*

Abstract. The WZ/ZZ (VZ) production channels with Z decaying to bottom quarks are one of the less reducible backgrounds for the standard model (SM) Higgs boson search in the same event topology (VH). The study of the VZ signal described here should be considered as a validation of the analysis strategy developed for the VH search in the boosted regime. The 8 TeV data sample including 19.0 fb^{-1} from the 2012 running period has been analyzed in the Z($\mu\mu$, ee, $\nu\nu$)Z and W($\mu\nu$, $e\nu$)Z channels and the presence of the signal has been established with a significance of 7.5σ.

1 Introduction

The search for the Higgs boson (H) decaying into a b-quark pair at the LHC is performed in events where H is produced in association with either a W or a Z boson (VH) decaying leptonically in order to suppress the multi-jet QCD background. A large boost for the $b\bar{b}$ pair and the V is also required to further suppress the V+jets and $t\bar{t}$ backgrounds.

The diboson production cross section is few times larger than the production cross section for VH [1]. Given the same event topology, VZ associated production is the least reducible background for the VH search, Fig. 1, left. The two channels can be separated only with very good di-jet mass resolution. For these reasons, the diboson process represents a standard candle to validate the Higgs boson search strategy.

In this note a study of the pp→VZ production mode is presented. The search is performed in a data sample corresponding to an integrated luminosity up to 19.0 fb^{-1} at $\sqrt{s}$ = 8 TeV, recorded by the CMS experiment at the LHC, in the W($\mu\nu$, $e\nu$,$\tau\nu$ [1])Z, Z(ee, $\mu\mu$,$\nu\nu$)Z channels.

This study is a validation of the VH($b\bar{b}$) analysis strategy developed by the CMS experiment, which reports a 2.1σ excess [2], and represents also the first attempt to measure the VZ cross section in the Z→ $b\bar{b}$ final state at CMS.

2 Event Selection

The event selection is based on the reconstruction of the leptonic decay of the V and of the Z boson decay into two b-tagged jets. Backgrounds are suppressed by requiring a large boost of the vector boson and the $b\bar{b}$ pair [3]. For each channel, different boost categories with different signal and background contributions are considered and the analysis is performed separately in each p_T(V) bin. Given the different p_T(V) dependence in VZ and VH processes, as shown in Fig. 1, right for ZH/ZZ, low p_T regions are more sensitive to the VZ production. Besides kinematics, also the back-to-back topology and minimal additional jet activity requirements are used to further suppress the background.

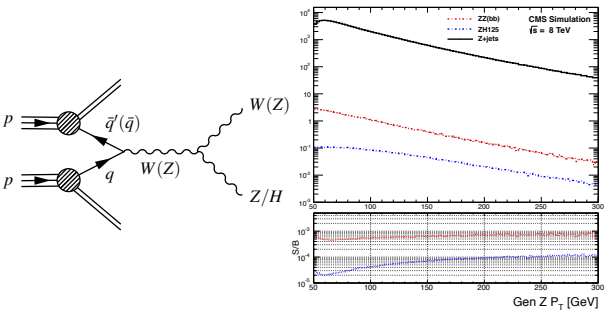

Figure 1. Left, Tree level s-channel Feynman diagram for VZ/VH production at LHC. Right, Differential distribution of the generated Z p_T for: Z+Jets, ZZ, ZH. The bottom plot contains the ratios ZZ/Z+jets and ZH/Z+jets.

W→ $\ell\nu$ decays are identified by requiring a single isolated lepton and additional E_T^{miss}. Z→ $\ell\ell$ candidates are reconstructed by combining isolated, oppositely charged pairs of electrons or muons and requiring the dilepton invariant mass to be within a Z mass window of 30 GeV width. The identification of Z→ $\nu\nu$ decays requires the

[a] e-mail: caterina.vernieri@cern.ch
[1] Only taus with 1-prong hadronic decays are explicitly considered.

E_T^{miss} in the event to be within the boost regions defined in Table 1.

The reconstruction of the $Z \to b\bar{b}$ decay is made by selecting the pair of jets in the event with the highest $p_T(jj)$. The two jets are tagged as b-jets using the Combined Secondary Vertex (CSV) algorithm [4].

3 Backgrounds Estimates

Control regions are identified in data and used to correct the Monte Carlo yields estimated for the main background processes: V+jets (light- and heavy-flavor) and $t\bar{t}$ production. A set of simultaneous fits is performed to several distributions of discriminating variables in the control regions, separately in each channel, to obtain consistent data/MC scale factors (SF).

Good agreement between data and simulation is found after applying the fitted SF in several control regions for all modes, as reported in [2].

4 Analysis Strategy

The main features of the Higgs search strategy, that this study aims to validate, are the use of the regression and multivariate analysis techniques, leading to an improvement of the resolution on the $b\bar{b}$ invariant mass and to a more efficient discrimination of the signal. More details on the analysis workflow are available in [2].

Signal and background yields are estimated in a fit region in the output discriminant of a boosted decision tree algorithm ("BDT anaysis") and as a cross check in the dijet invariant mass distribution ("M($b\bar{b}$) analysis"), using shape analysis techniques. The first method, by making use of correlations between discriminating variables in signal and background events, yields a new variable that allows to discriminate the signal more effectively than with the use of the M($b\bar{b}$) information only.

4.1 b-jet Energy Regression

The Z boson mass resolution is improved by applying regression techniques similar to those used by the CDF experiment [5]. The b-jet energy resolution is worse with respect to the light quark/gluon induced jets, since in about the 20% of cases a neutrino is present in the B hadron decay. The regression technique allows to improve the energy resolution of b-jets and the mass resolution of the $b\bar{b}$ pair improves as well.

The corrections to the jet energy are obtained through a BDT discriminant, trained with inputs that include detailed information about the jet structure and basic kinematic information, b-tag and soft lepton information when available. For the Z($\ell\ell$) channel also the information carried by the variables related to the E_T^{miss} vector is exploited, since there is no real E_T^{miss} in the event.

The improvement on the mass resolution is approximately 15% and results in a better separation of the VZ/VH processes as shown in Fig. 2.

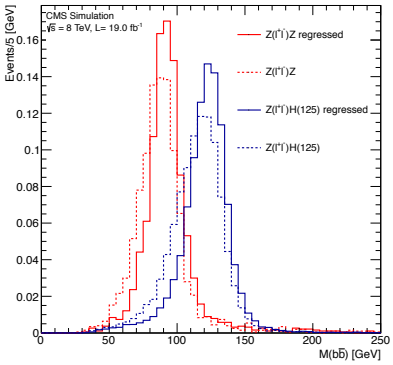

Figure 2. Mass difference between ZH and ZZ simulated processed before and after the regression is applied.

Table 1. Selection criteria for the samples used in the BDT training in each channel. The values listed for kinematical variables are in units of GeV.

Variable	W($\ell\nu$)Z	Z($\ell\ell$)Z	Z($\nu\nu$)Z
$m_{\ell\ell}$	–	[75 – 105]	–
$p_T(j_1)$, $p_T(j_2)$	> 30, 30	> 20, 20	> 60, 30
p_T(jj)	> 100 (e, μ) > 120 (τ)	–	> 100
M(jj)	< 250	[40 – 250]	< 250
$p_T(\ell)$	> 30 (e, μ) > 40 (τ)	> 20	–
	[100-130] (μ)	[50-100]	[100-130]
p_T(V)	[100-150] (e) [130-180] (μ)	–	[130-170]
	> 150, 180, 120 (e,μ,τ)	–	> 170
CSV_{max}, CSV_{min}	> 0.40,0.40	> 0.50,0.24	>0.67 (0.24)
N_{aj}, N_{al}	– , = 0	– , –	< 2, = 0
$\Delta\phi$(V, H)	–	–	> 2.0
$\Delta\phi(E_T^{miss}$, jet)	–	–	> 0.7 (0.7,0.5)
$\Delta\phi(E_T^{miss}$, trkMET)	–	–	< 0.5
$\Delta\phi(E_T^{miss}$, lep)	<π/ 2	–	–

4.2 BDT Analysis

As a validation of the multivariate technique, BDT discriminants are trained using the diboson sample as signal, and all other processes, including VH production (at the predicted SM rate for a 125 GeV Higgs boson), as background. The event selection used in the BDT analysis is reported in Table 1. The set of input variables used is chosen by iterative optimization from a larger number of potentially discriminating variables (Table 2). The signal and background yields are then estimated in a fit region defined in the continuous output of the BDT, and a shape analysis is performed on that output.

In the Z($\nu\nu$)Z and W($\ell\nu$)Z channels a 5-10% improvement in the analysis sensitivity, is reached using a background-specific approach for the BDT ("mBDT"). Several BDTs are trained to discriminate the signal from specific backgrounds. Events are then classified based on the values of the discriminant variables of each BDT. This technique, similar to the one used by the CDF collaboration in [7], divides the samples into distinct subsets that are enriched in $t\bar{t}$, V+jets and diboson events. Two background-specific BDT discriminants are trained to separate $t\bar{t}$ and V+jets from VZ. The output distributions of the background-specific BDTs are then used to separate events in three categories. Those that fail a cut on the $t\bar{t}$

Table 2. Variables used in the BDT training.

Variable
p_T(j): transverse momentum of each Z($b\bar{b}$) daughter
M(jj): dijet invariant mass
p_T(jj): dijet transverse momentum
p_T(V): vector boson transverse momentum (or E_T^{miss})
CSV_{max}: value of CSV for the Z($b\bar{b}$) daughter with largest CSV value
CSV_{min}: value of CSV for the Z($b\bar{b}$) daughter with second largest CSV
$\Delta\phi$(V, H): azimuthal angle between V and dijet
$\Delta\eta$(jj): difference in η between Z($b\bar{b}$) daughters
ΔR(jj): distance in η–ϕ between Z($b\bar{b}$) daughters
N_{aj}: number of additional jets
$\Delta\theta_{pull}$: color pull angle [6]
$\Delta\phi(E_T^{miss}$, jet): azimuthal angle between $E_T^{miss}$and the closest jet
$maxCSV_{aj}$: maximum CSV of the additional jets in an event
minΔR(H, aj): minimum distance between an additional jet and Z($b\bar{b}$)
Angular variables: VZ mass, Angle Z-Z*, Angle Z-l, Angle Z-jet (Z($\ell\ell$) only)

BDT are classified as $t\bar{t}$-like events, those that pass the $t\bar{t}$ BDT cut but fail a cut on the V+jets BDT are classified as V+jets-like events. Finally, those that pass all BDT cuts are VZ-enriched events. The events in each category are then processed by the final BDT discriminant and the resulting distribution, now composed of three distinct subsets of events, is used as input to the fitting program. An example of the output distribution is reported in Fig.3.

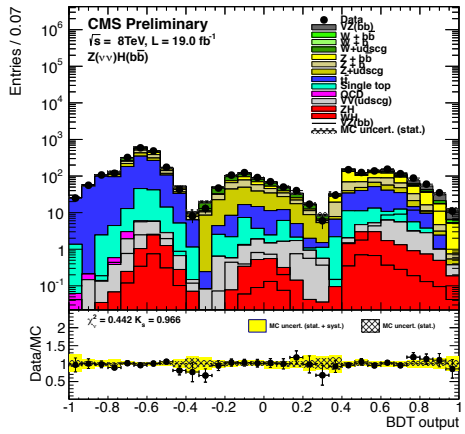

Figure 3. BDT output distribution for Z($\nu\nu$) in the high p_T(V) bin, for data (points with errors), all backgrounds, and signal.

4.3 M($b\bar{b}$) cross check

As a cross-check to the multivariate analysis, a simpler analysis is done by performing a fit to the shape of the dijet invariant mass distribution M($b\bar{b}$). The event selection for this analysis is more stringent than the one used in the BDT analysis [2]. Fig. 4 shows the weighted dijet invariant mass distribution for the combination of all six channels, in all p_T(V) bins, in the combined data samples corresponding to integrated luminosities of 5.0 fb^{-1} at $\sqrt{s}$ =7 TeV and up to 19.0^{-1} at $\sqrt{s}$ =8 TeV, used in the analogue Higgs search. The data are consistent with the presence of a diboson signal (ZZ and WZ, with Z→ $b\bar{b}$), with a rate consistent with the SM prediction evaluated with a simulation based on the MADGRAPH generator, together with a small excess consistent with originating from the production of a 125 GeV SM Higgs boson.

5 Results

Performing the mBDT analysis with the specific-background approach, the observed excess of events for the combined WZ and ZZ with $Z \rightarrow b\bar{b}$ processes has a p-value of 6.4×10^{-14} corresponding to 7.5 standard deviations. The signal strength, relative to the prediction using the diboson MADGRAPH generator, is $1.19^{+0.28}_{-0.30}$. In Table 3 the results for the M($b\bar{b}$) analysis and for the BDT approach are reported.

Table 3. Expected and observed significances, the best-fit signal strength modifier μ of the excess of events above the estimated background on the diboson production in $b\bar{b}$ final state.

Mode	M($b\bar{b}$)	mBDT
Combined Exp.	4.3 σ	6.3 σ
Combined Obs.	3.7 σ	7.5 σ
Signal strength μ	$0.78^{+0.26}_{-0.28}$	$1.19^{+0.28}_{-0.3}$

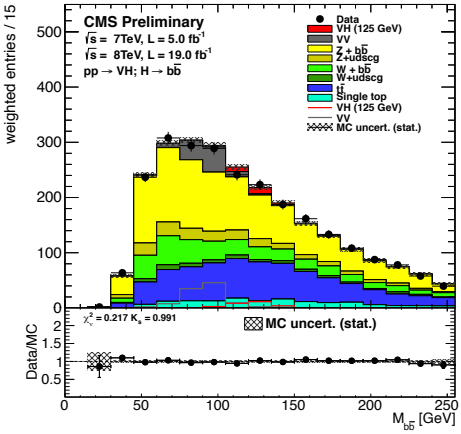

Figure 4. Weighted dijet invariant mass distribution, combined for all channels. For each channel, the relative weight of each p_T(V) bin is obtained from the ratio of the expected number of signal events to the sum of expected signal and background events.

References

[1] LHC Higgs Cross Section Working Group, **CERN-2011-002**, (2011).

[2] CMS Collaboration, Physics Analysis Summary **HIG-13-012**, (2013).

[3] J. M. Butterworth et al., Phys. Rev. Lett. **100**, 242001 (2008).

[4] CMS Collaboration, JINST **8**, P04013 (2013).

[5] T. Aaltonen et al., arXiv:1107.3026 (2011).

[6] J. Gallicchio and M. D. Schwartz, Phys. Rev. Lett. **105**, 022001 (2010).

[7] CDF Collaboration, Phys. Rev. Lett. **109**, 111803 (2012).